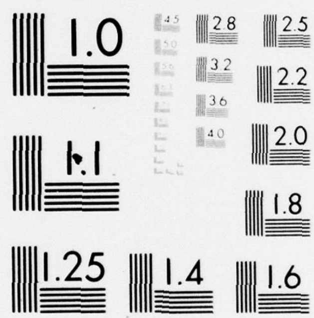
AD-A045 943 DELAWARE UNIV NEWARK DEPT OF STATISTICS AND COMPUTER--ETC F/G 12/2
THE M/M/1 QUEUE WITH RANDOMLY VARYING ARRIVAL AND SERVICE RATES--ETC(U)
JUL 77 M F NEUTS AFOSR-77-3236
UNCLASSIFIED TR-77/11 AFOSR-TR-77-1233 NL

1 OF 1
ADA045943

END
DATE
FILMED
12-77
DDC


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0

AFOSR-TR- 77-1233

6

B.S.

AD A045943

Department of

STATISTICS AND COMPUTER SCIENCE

UNIVERSITY OF DELAWARE
Newark, Delaware 19711

The M/M/1 Queue with Randomly Varying Arrival and Service Rates

by

Marcel F. Neuts

University of Delaware

Department of Statistics
and
Computer Science
Technical Report No. 77/11

July 1977

This research was sponsored by the Air Force Office of Scientific Research Air Force Systems Command USAF, under Grant No. AFOSR-77-3236. The United States Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright notation hereon.

see 1473

ABSTRACT

We study computationally feasible solutions for a number of problems, related to a M/M/1 queue in which the arrival and service rates vary according to the state of an underlying Markov chain.

Our results may be used to model the effect of rush-hour phenomena or other extraneous fluctuations on the characteristics of an M/M/1 queue.

KEY WORDS

M/M/1 queue, random environment, *queue length*, *waiting time*, quasi-birth-and-death processes, Markov chains, nonlinear equations, computational probability

I. Introduction

Consider an m-state, irreducible, continuous-parameter Markov chain with infinitesimal generator Q, which describes a randomly varying "environment" for a queue of M/M/1 type. Specifically we assume that whenever the Markov chain is in the state j, there is an arrival rate λ_j to a single-server queue and a service rate μ_j, with $\lambda_j>0$, $\mu_j>0$, $1\le j\le m$. When the state of the Markov chain changes, so do both the arrival and service rates. This model was introduced by U. Yechiali and P. Naor [8] and further investigated by U. Yechiali [9] and P. Purdue [7]. It provides a tractable description of a simple queue, subject to rush-hour behavior or other extraneous phase fluctuations.

In this paper, we solve the M/M/1 queue in a random environment by an approach, which leads to easily implementable algorithms for the numerical computation of the relevant stationary distributions.

By $\underline{\lambda}$ and $\underline{\mu}$, we denote the m-vectors with components λ_j and μ_j, $1\le j\le m$, respectively. For any vector $\underline{a}$, we introduce the matrix $\Delta(\underline{a})=\mathrm{diag}(a_1,\ldots,a_m)$. The matrices A_0, A_1 and A_2 are defined by $A_0=\Delta(\underline{\mu})$, $A_1=Q-\Delta(\underline{\lambda}+\underline{\mu})$, $A_2=\Delta(\underline{\lambda})$. The invariant probability vector of the matrix Q is denoted by $\underline{\pi}$ and is the unique solution of the system $\underline{\pi}Q=\underline{0}$, with $\underline{\pi}\underline{e}=1$, where $\underline{e}=(1,1,\ldots,1)'$.

The queueing model of interest is then described by a continuous-parameter Markov chain on the state space $\{(i,j),\ i\ge 0,\ 1\le j\le m\}$. The chain is in the state (i,j), when i customers are present in the system and the Q-process is in the state j. The infinitesimal generator Q* of the chain is given by

$$(1) \qquad Q^* = \begin{vmatrix} A_0+A_1 & A_2 & 0 & 0 & \cdots \\ A_0 & A_1 & A_2 & 0 & \cdots \\ 0 & A_0 & A_1 & A_2 & \cdots \\ 0 & 0 & A_0 & A_1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \end{vmatrix},$$

and is of a form, studied by V. Wallace [10] under the name of quasi-birth-and-death processes. We shall show that the invariant probability vector $\underline{x}$ of the matrix Q^*, if it exists, is of a matrix-geometric form and may easily be computed. Before doing so, we discuss a number of other points of independent interest.

Lemma 1

The inverse A_1^{-1} exists and is strictly negative. The matrices $C_0=-A_1^{-1}A_0$, $C_2=-A_1^{-1}A_2$, $B_0=-A_2A_1^{-1}$ and $B_2=-A_0A_1^{-1}$ are strictly positive. The matrix $B=B_0+B_2$ has a spectral radius equal to one. The matrix $C=C_0+C_2$ is stochastic.

The vectors $\underline{\pi}$ and $\underline{v}=(\underline{\pi}A_1\underline{e})^{-1}A_1\underline{e}$ are respectively positive left and right invariant vectors of B and $\underline{\pi}\underline{v}=1$.

The vectors $\underline{u}=(\underline{\pi}A_1\underline{e})^{-1}\underline{\pi}A_1$ and $\underline{e}$ are respectively positive left and right invariant vectors of the matrix C and $\underline{u}\underline{e}=1$.

The inequalities $\underline{u}(2C_2\underline{e})\leq 1$ and $\underline{\pi}(2B_2\underline{v})\geq 1$, are each equivalent to $\rho=(\underline{\pi}\underline{\lambda})(\underline{\pi}\underline{\mu})^{-1}\leq 1$.

Proof

Since the matrix Q is irreducible, the matrix

$$(2) \qquad -A_1^{-1} = -[Q-\Delta(\underline{\lambda}+\underline{\mu})]^{-1} = \int_0^\infty \exp[Q-\Delta(\underline{\lambda}+\underline{\mu})]t\, dt,$$

is strictly positive [1]. The positivity of B_0, B_2, C_0 and C_2 is

now obvious.

Since $\underline{\pi}(A_0+A_1+A_2)=\underline{\pi}Q=\underline{0}$, and $(A_0+A_1+A_2)\underline{e}=Q\underline{e}=\underline{0}$, it readily follows that

$$(3) \quad \underline{\pi}B=\underline{\pi}, \qquad BA_1\underline{e}=A_1\underline{e},$$
$$\underline{\pi}A_1C=\underline{\pi}A_1, \qquad C\underline{e}=\underline{e}.$$

Since the vector $\underline{\pi}$ is positive, the first equality in (3) shows that the spectral radius of B is one.

The inner products $\underline{u}(2C_2\underline{e})$ and $\underline{\pi}(2B_2\underline{v})$ are given by

$$(4) \quad 2\ \underline{u}\ C_2\ \underline{e} = -2(\underline{\pi}A_2\underline{e})(\underline{\pi}A_1\underline{e})^{-1} = 2\ \underline{\pi}\ \underline{\lambda}[\underline{\pi}\underline{\lambda}+\underline{\pi}\underline{\mu}]^{-1},$$
$$2\ \underline{\pi}\ B_2\ \underline{v} = -2(\underline{\pi}A_0\underline{e})(\underline{\pi}A_1\underline{e})^{-1} = 2\ \underline{\pi}\ \underline{\mu}[\underline{\pi}\underline{\lambda}+\underline{\pi}\underline{\mu}]^{-1},$$

so that the stated inequalities are each equivalent to $\rho\leq 1$.

II. The Busy Period

We consider the queue, starting in the state $(i+1,j)$ at time $t=0$, and examine the first passage time to the set of states $\underline{i}=\{(i,j'),\ 1\leq j'\leq m\}$. This first passage time corresponds to the familiar busy period in simple queues.

By $\tilde{G}_{jj'}(k,x)$, $k\geq 1$, $x\geq 0$, $1\leq j$, $j'\leq m$, we denote the probability that, starting in the state $(i+1,j)$, the first visit to the set $\underline{i}$ occurs no later than time x, into the state (i,j') and exactly k service completions occur during the first passage time.

For convenience, we introduce the transforms

$$(5) \quad G^*_{jj'}(z,s) = \sum_{k=1}^{\infty} z^k \int_0^{\infty} e^{-sx}\ d\ \tilde{G}_{jj'}(k,x),$$

and the matrix $G^*(z,s)=\{G^*_{jj'}(z,s)\}$.

The first passage problem under consideration is of a type, that was extensively examined by the author. We shall only present the essential points here and refer for the detailed proofs to [3] and [5].

Theorem 1

The matrix $G^*(z,s)$ satisfies the equation

$$(6) \quad G^*(z,s) = z(sI-A_1)^{-1}A_0 + (sI-A_1)^{-1}A_2\, G^{*2}(z,s),$$

for $s \geq 0$, $0 \leq z \leq 1$. In an appropriately defined set of transform matrices, $G^*(z,s)$ is the unique solution to (6).

The queue is stable if and only if the matrix $G=G^*(1,0)$ is stochastic. The matrix G is the minimal solution in the set of substochastic matrices to the equation

$$(7) \quad G = C_0 + C_2G^2.$$

The matrix G is stochastic if and only if $\rho \leq 1$ and is unique and strictly positive.

Proof

Equation (6) follows from a standard first passage argument by considering the first time that the queue length goes either down or up. The other statements were proved in [3], where it is also shown that G is stochastic if and only if the inequality $\underline{u}(2C_2\underline{e}) \leq 1$ holds. From Lemma 1, we know that the latter is equivalent to $\rho \leq 1$. This is also the equilibrium condition obtained by U. Yechiali [9]. Equation (6) was also derived and discussed by P. Purdue [7].

In the remainder of the paper, we assume that $\rho \leq 1$. The matrix G may be computed by successive substitutions in Equation (7). We

shall denote the invariant probability vector of G by $\underline{g}$ and by $\tilde{G}$ an m×m matrix with identical rows given by $\underline{g}$.

The following theorem gives explicit expressions for the expected duration of and for the mean number of customers served during a busy period.

We define the vectors $\underline{\mu}^*$ and $\underline{\mu}^\circ$ by

$$(8) \qquad \underline{\mu}^* = -\left[\frac{\partial}{\partial s} G^*(z,s)\underline{e}\right]_{\substack{z=1\\s=0}}, \qquad \underline{\mu}^\circ = \left[\frac{\partial}{\partial z} G^*(z,s)\right]_{\substack{z=1\\s=0}}.$$

The quantity μ_j^* is then the expected duration of a busy period, starting with one customer and with the Q-process in the state j. The quantity μ_j° is the expected number of departures during such a busy period.

Theorem 2

If $\rho<1$,

$$(9) \qquad \underline{\mu}^* = -(I-G+\tilde{G})[Q+\Delta(\underline{\lambda}-\underline{\mu})\tilde{G}]^{-1}\underline{e},$$

$$\underline{\mu}^\circ = -(I-G+\tilde{G})[Q+\Delta(\underline{\lambda}-\underline{\mu})\tilde{G}]^{-1}\underline{\mu},$$

and

$$(10) \qquad \underline{g}\underline{\mu}^* = (\underline{\pi}\underline{\mu})^{-1}(1-\rho)^{-1}, \qquad \underline{g}\underline{\mu}^\circ = (1-\rho)^{-1}.$$

If $\rho=1$, the vectors $\underline{\mu}^*$ and $\underline{\mu}^\circ$ are infinite.

Proof

The formulas (9) can be obtained by particularizing results in [5], but as the proof is short, we repeat the essential steps. By routine differentiations in (6), we obtain

$$(11)\quad (A_1+A_2+A_2G)\underline{\mu}^*=-\underline{e},$$
$$(A_1+A_2+A_2G)\underline{\mu}^\circ=-A_0\underline{e}=-\underline{\mu}.$$

Since $I-G+\tilde{G}$ is nonsingular and since also $A_0+A_1G+A_2G^2=0$, we have that

$$(12)\quad (A_1+A_2+A_2G)(I-G+\tilde{G})=A_0+A_1+A_2+(A_1+2A_2)\tilde{G}=Q+\Delta(\underline{\lambda}-\underline{\mu})\tilde{G},$$

which yields the formulas (9). In [5], it is shown that the matrix in (12) is nonsingular if $\rho<1$ and becomes singular for $\rho=1$.

Finally, the formulas (10) follow by noting that

$$(13)\quad \underline{g}=\underline{g}(I-G+\tilde{G}),$$
$$\underline{\pi}[Q+\Delta(\underline{\lambda}-\underline{\mu})\tilde{G}]=(\underline{\pi}\underline{\lambda}-\underline{\pi}\underline{\mu})\underline{g}.$$

The formulas (10) provide powerful accuracy checks in numerical computations.

Corollary 1

The transform matrix corresponding to a first passage from the set of states $\underline{i+r}$ to the set of states $\underline{i}$ is given, for $r\geq 1$, by $[G^*(z,s)]^r$. The expected duration of and the mean number of customers initially and with the Q-chain in the state j, are given respectively by the j-th components of the vectors

$$(14)\quad \underline{\mu}^*(r)=-(I-G^r+r\tilde{G})[Q+\Delta(\underline{\lambda}-\underline{\mu})\tilde{G}]^{-1}\underline{e},$$
$$\underline{\mu}^\circ(r)=-(I-G^r+r\tilde{G})[Q+\Delta(\underline{\lambda}-\underline{\mu})\tilde{G}]^{-1}\underline{\mu}.$$

Proof

The first statement follows directly by probabilistic considerations [3] and by standard differentiations we obtain

$$(15)\quad \underline{\mu}^*(r)=\sum_{\nu=0}^{r-1} G^{\nu}\underline{\mu}^*=(I-G^r+r\tilde{G})(I-G+\tilde{G})^{-1}\underline{\mu}^*,$$

and similarly for $\underline{\mu}^{\circ}(r)$.

III. The Effective Service and Interarrival Times

In this section, we consider the probability distribution of a service time starting at time $t=0$, with the Q-process in the state i. This will be called the effective service time starting in state i. The results for the effective interarrival times are similar and will be stated without proofs.

Let $\psi_{ij}(\nu,t)$, $\nu\geq 0$, $t\geq 0$, $1\leq i,j\leq m$, be the probability that a service, starting at time 0 in the state i, lasts for a time t at least and that during $(0,t]$, there are $\nu\geq 0$ new arrivals to the queue. A direct birth-and-death argument yields

$$(16)\quad \psi'_{ij}(\nu,t)=\delta_{ij}(-\lambda_i-\mu_i+Q_{ii})\psi_{ii}(\nu,t)+\sum_{h\neq i} Q_{ih}\psi_{hj}(\nu,t)$$
$$+(1-\delta_{\nu 0})\lambda_i\ \psi_{ij}(\nu-1,t),$$

for $t\geq 0$, $\nu\geq 0$, $1\leq i,j\leq m$, with initial conditions $\psi_{ij}(\nu,0)=\delta_{\nu 0}\delta_{ij}$, for $\nu\geq 0$. By δ_{ij}, we denote the usual Kronecker delta.

In matrix notation we obtain the recursive system of differential equations

$$(17)\quad \psi'(0,t)=A_1\psi(0,t),$$
$$\psi'(\nu,t)=A_1\psi(\nu,t)+A_2\psi(\nu-1,t),\qquad \text{for } \nu\geq 1.$$

This readily leads to

$$(18)\quad \psi^*(z,t)=\sum_{\nu=0}^{\infty} \psi(\nu,t)z^{\nu}=\exp[(A_1+zA_2)t],\qquad \text{for } 0\leq z\leq 1,\quad t\geq 0.$$

Formula (18) has a number of useful consequences, which we combine into the following theorem.

Theorem 3

The probability that a service, starting at time 0 in the state i ends during (t,t+dt] with the Q-process in the state j, is given by the (i,j)-th entry of the matrix

$$(19)\quad \exp[(A_1+A_2)t]A_0dt=\exp\{[Q-\Delta(\underline{\mu})]t\}\Delta(\underline{\mu})dt.$$

For any initial probability vector $\underline{\gamma}$ over the states 1,...,m of the Q-process, the distribution of the effective service time is a distribution of phase type [4] with the representation $[\underline{\gamma},\ Q-\Delta(\underline{\mu})]$. Its mean E_s is given by

$$(20)\quad E_s=\underline{\gamma}[\Delta(\underline{\mu})-Q]^{-1}\underline{e}.$$

The probability generating function $p_i(z)$ of the number of arrivals during a service starting in the state i is given by the i-th component of the vector

$$(21)\quad \underline{p}(z)=\int_0^\infty \exp[(A_1+zA_2)t]A_0\underline{e}dt$$

$$=[(1-z)\Delta(\underline{\lambda})+\Delta(\underline{\mu})-Q]^{-1}\underline{\mu}.$$

The matrix $[\Delta(\underline{\mu})-Q]^{-1}\Delta(\underline{\mu})$ is stochastic and strictly positive. Its invariant probability vector $\underline{\pi}^*$ is given by

$$(22)\quad \underline{\pi}^*=(\underline{\pi}\underline{\mu})^{-1}\underline{\pi}\Delta(\underline{\mu}).$$

If the state i of the Q-chain is chosen according to the vector $\underline{\pi}^*$, the corresponding service time will be called <u>the average effective service time</u>. Its mean E_s^* is given by $(\underline{\pi}\underline{\mu})^{-1}$ and the

average number of arrivals during it is given by $\underline{\pi}^*\underline{p}'(1)=\rho$.

Proof

The first statement follows immediately from (18). The effective service time with the initial probability vector $\underline{\gamma}$ has the same probability distribution as that of the time till absorption in the Markov chain with infinitesimal generator

$$\begin{vmatrix} Q-\Delta(\underline{\mu}) & \underline{\mu} \\ \underline{0} & 0 \end{vmatrix},$$

and initial probability vector $(\underline{\gamma},0)$. It is therefore a PH-distribution and may easily be computed numerically. The expression for the mean E_s is immediate. [4]

The expression for $\underline{p}(z)$ follows directly from (18). Since $\Delta(\underline{\mu})\underline{e}-Q\underline{e}=\underline{\mu}$, it follows that the matrix $[\Delta(\underline{\mu})-Q]^{-1}\Delta(\underline{\mu})$ is stochastic. That $\underline{\pi}^*$ is its invariant vector may be directly verified.

From (21), we obtain upon differentiation that

(23) $\underline{p}'(1)=[\Delta(\underline{\mu})-Q]^{-1}\underline{\lambda}$,

so that

(24) $\underline{\pi}^*\underline{p}'(1)=(\underline{\pi}\underline{\mu})^{-1}\underline{\pi}\ \Delta(\underline{\mu})[\Delta(\underline{\mu})-Q]^{-1}\Delta(\underline{\mu})\ \Delta^{-1}(\underline{\mu})\underline{\lambda}=(\underline{\pi}\underline{\mu})^{-1}(\underline{\pi}\underline{\lambda})=\rho$.

A similar calculation yields that $E_s^*=(\underline{\pi}\underline{\mu})^{-1}$. The interpretation of ρ implied by (24) was first pointed out in [7].

The corresponding result for the interarrival times are as follows.

Theorem 4

Given that an arrival occurs at t=0 and that the Q-process is in the state i, the probability that the next arrival occurs during (t,t+dt] with the Q-process in the state j, is the (i,j)-entry of

the matrix

(25) $\exp[(A_0+A_1)t]A_2dt=\exp[(Q-\Delta(\underline{\lambda}))t]\Delta(\underline{\lambda})dt.$

For any initial probability vector $\underline{\gamma}$ over the states $1,\ldots,m$, the effective interarrival time has a PH-distribution with representation $[\underline{\gamma},Q-\Delta(\underline{\lambda})]$ and mean $E_T=\underline{\gamma}[\Delta(\underline{\lambda})-Q]^{-1}\underline{e}$.

Given an infinite supply of customers at t=0, the probability generating function of the number of departures during an interarrival interval starting in the state i, is given by

(26) $\tilde{\underline{p}}(z)=[\Delta(\underline{\lambda})+(1-z)\Delta(\underline{\mu})-Q]^{-1}\underline{\lambda}.$

For $\underline{\gamma}=\tilde{\underline{\pi}}=(\underline{\pi}\underline{\lambda})^{-1}\underline{\pi}\Delta(\underline{\lambda})$, we obtain the average effective interarrival time and the mean number of departures during the average effective interarrival time is given by ρ^{-1} and the mean duration of the latter is $(\underline{\pi}\underline{\lambda})^{-1}$.

IV. The Steady-state Queue Length

This section is devoted to the proof of the following statements.

Theorem 5

If $\rho<1$, the invariant probability vector $\underline{x}$ of the Markov chain with infinitesimal generator Q^* is given by $\underline{x}=(\underline{x}_0,\underline{x}_1,\ldots)$, where

(27) $\underline{x}_k=\underline{\pi}(I-R)R^k,$ for $k\geq 0$.

The matrix R is the unique solution in the set of nonnegative matrices of order m, which have a spectral radius less than one, of the equation

(28) $R^2A_0+RA_1+A_2=0.$

The matrix R is strictly positive and $\underline{\pi}R<\underline{\pi}$.

Proof

The invariant vector will be of the stated form, if there exists a matrix R with the stated properties, such that (28) holds and there exists a vector $\underline{x}_0>\underline{0}$, such that

$$(29) \quad \underline{x}_0(A_0+A_1+RA_0)=\underline{0}.$$

We first show that $\underline{x}_0=\underline{\pi}(I-R)$ and we shall verify below that $\underline{x}_0$ is strictly positive. Equations (28) and (29) yield

$$(30) \quad \underline{x}_0(A_0+A_1+RA_0)+\sum_{\nu=0}^{\infty} \underline{x}_0R^{\nu}(R^2A_0+RA_1+A_2)=$$

$$\underline{x}_0(I-R)^{-1}(A_0+A_1+A_2)=\underline{x}_0(I-R)^{-1}Q=\underline{0}.$$

Since also $\underline{x}_0(I-R)^{-1}\underline{e}=1$, (30) implies that $\underline{x}_0=\underline{\pi}(I-R)$.

The equation (28) may be written as

$$(31) \quad R=R^2B_2+B_0.$$

Let {R(n)} be the sequence of matrices obtained from successive substitutions, starting with R(0)=0, in (31). As was done in [6], one may then verify that

$$(32) \quad R(n+1)\geq R(n), \qquad \underline{\pi}R(n)\leq\underline{\pi},$$

so that the spectral radius $sp[R(n)]\leq 1$. The matrices R(n) therefore converge to a matrix R, which is strictly positive, has $sp(R)\leq 1$, and satisfies (31). That matrix is also the minimal nonnegative solution to Equation (31).

By repeating verbatim the argument given in [6], Lemma 4, the spectral radius η of R is the smallest positive solution of the

equation

(33) $z=\chi(z)$, $\qquad 0\le z\le 1$.

where $\chi(z)$ is the Perron eigenvalue of the positive matrix $B_2z^2+B_0$.

Setting $z=e^{-s}$, Equation (33) may be written as

(34) $s=-\log\chi(e^{-s})$, $\qquad s\ge 0$.

A theorem of J.F.C. Kingman [2] guarantees that the function $\log\chi(e^{-s})$ is convex for $s\ge 0$. It is also clearly decreasing, negative for $s>0$ and tends to the finite limit $\log sp(B_0)$ as $s\to\infty$. The equation (34) has the solution $s=0$, since $sp(B)=1$. There is a unique positive solution $s_0=-\log\eta$, if and only if $\chi'(1-)>1$.

A direct calculation, similar to that presented in [6], Lemma 4, yields that

(35) $\chi'(1-)=\underline{\pi}(2B_2)\underline{v}$,

where $\underline{v}$ is the right invariant vector of B, introduced in Lemma 1. It follows that $\chi'(1-)>1$, if and only if $\rho<1$.

So, provided that $\rho<1$, the matrix R has spectral radius less than one. The uniqueness of the solution R is proved exactly as in [6].

It remains to show that $\underline{\pi}R<\underline{\pi}$. From (32), we have $\underline{\pi}R\le\underline{\pi}$. The equations $\underline{\pi}(A_0+A_1+A_2)=\underline{0}$ and $\underline{\pi}(R^2A_0+RA_1+A_2)=\underline{0}$ imply that

(36) $\underline{\pi}-\underline{\pi}R=-(\underline{\pi}-\underline{\pi}R^2)A_0A_1^{-1}$.

Since $\underline{\pi} \geq \underline{\pi}R^2$, but $\underline{\pi} \neq \underline{\pi}R^2$ and the matrix $-A_0A_1^{-1}$ is strictly positive, it is clear that the vector $\underline{\pi}(I-R)$ is strictly positive.

Remarks

Theorem 5 has a large number of straightforward, but useful consequences. Once the easily computed matrix R is known, all moments, marginal and conditional densities of the queue length are known.

The conditional densities

$$(37)\qquad q_i(j)=\pi_j^{-1}[\underline{\pi}(I-R)R^i]_j, \qquad i\geq 0,$$

of the queue length, given that the Q-process is in the state j, shed light on the oscillatory behavior of the queue length in the steady-state, at least for such choices of the parameters which correspond to alternating periods of high and low traffic.

V. The Steady-state Virtual Waiting Time

Assume the queue in steady-state at time 0. Let $W_j(x)$ be the probability that a (virtual) customer arriving at that time will enter service no later than time x and that the Q-process will be in the state j at the beginning of his service.

It is easy to see that $W_j(x)$ is also the probability that in the Markov chain with infinitesimal generator Q_w, given by

$$(38)\qquad Q_w = \begin{vmatrix} 0 & 0 & 0 & \cdots \\ \Delta(\underline{\mu}) & Q-\Delta(\underline{\mu}) & 0 & \cdots \\ 0 & \Delta(\underline{\mu}) & Q-\Delta(\underline{\mu}) & \cdots \\ 0 & 0 & \Delta(\underline{\mu}) & \cdots \\ 0 & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \end{vmatrix},$$

and initial probability vector $\underline{x}=(\underline{x}_0,\underline{x}_1,\ldots)$ with $\underline{x}_k=\underline{\pi}(I-R)R^k$, $k\geq 0$, absorption into the set of states $\underline{0}=\{(0,1),\ldots,(0,m)\}$ occurs no later than time x into the state (0,j).

The vector $\underline{W}(x)$ with components $W_j(x)$, $1\leq j\leq m$, is, in general, not expressible in a closed form. The vector $\underline{W}^*(s)$ of the Laplace-Stieltjes transforms of $\underline{W}(\cdot)$ is given by

$$(39)\quad \underline{W}^*(s)=\sum_{k=0}^{\infty}\underline{\pi}(I-R)R^k\{[sI+\Delta(\underline{\mu})-Q]^{-1}\Delta(\underline{\mu})\}^k,$$

for Re $s\geq 0$.

The <u>time-in-system</u> of the virtual customer arriving at time 0 can be studied in the same manner. Let $\tilde{W}_j(x)$, with Laplace-Stieltjes transform $\tilde{W}_j^*(s)$, be the probability that a virtual customer arriving at time 0, <u>leaves</u> the system (under the first-come, first-served discipline) no later than time x with the Q-process in the state j at the time of his departure.

By using the results obtained for the distribution of the effective service time, one immediately obtains that

$$(40)\quad \underline{\tilde{W}}^*(s)=\underline{W}^*(s)[sI+\Delta(\underline{\mu})-Q]^{-1}\Delta(\underline{\mu}).$$

Computation of $\underline{W}(\cdot)$ and $\underline{\tilde{W}}(\cdot)$

Although $\underline{W}(\cdot)$ and $\underline{\tilde{W}}(\cdot)$ are not tractable in a convenient analytic manner, they can easily be computed as follows.

For $\underline{W}(\cdot)$, we form the infinite system of differential equations

$$(41)\quad \underline{y}_k'(x)=\underline{y}_k(x)[Q-\Delta(\underline{\mu})]+\underline{y}_{k+1}(x)\Delta(\underline{\mu}),$$

for $k\geq 1$, $x\geq 0$, with the initial conditions

(42) $\underline{y}_k(0)=\underline{\pi}(I-R)R^k$, $\qquad k\geq 1$.

For every $x\geq 0$, the vector $\underline{W}(x)$ is then given by

(43) $\underline{W}(x)=\underline{\pi}(I-R)+\int_0^x \underline{y}_1(u)du\Delta(\underline{\mu})$,

and

(44) $\underline{W}(x)\underline{e}=1-\sum_{k=1}^{\infty} \underline{y}_k(x)\underline{e}$.

Remarks

a. If only $\underline{W}(x)\underline{e}$ and not $\underline{W}(x)$ is to be computed, there is a slight gain in efficiency by solving for the vectors $\underline{Y}_k(x)=\sum_{\nu=k}^{\infty} \underline{y}_\nu(x)$, after modifying the system (41) in the obvious manner. The summation in (44) is then eliminated.

b. The Markov chain Q_W can only move towards lower states. It is therefore obvious how to truncate the system of differential equations (41). In order to lose at most a probability mass ε in tail of the probability distribution $\underline{W}(x)\underline{e}$, one truncates at the index K such that

(45) $\underline{\pi}R^{K+1}\underline{e}=\sum_{\nu=K+1}^{\infty} \underline{\pi}(I-R)R^{\nu}\underline{e}<\varepsilon$.

This bounds the error due to truncation of the infinite system. The global error involved in solving the resulting finite system of differential equations needs to be considered separately.

c. The vector $\underline{W}(\cdot)$ can be evaluated by solving the system (41) with the initial conditions

(46) $\underline{y}_k(0)=\underline{\pi}(I-R)R^{k-1}$, $\qquad$ for $k\geq 1$.

Formula (40) also leads to

(47) $\underline{\tilde{W}}'(x)+\underline{\tilde{W}}(x)\Delta^{-1}(\underline{\mu})[Q-\Delta(\underline{\mu})]\Delta(\underline{\mu})=\underline{W}'(x)\Delta(\underline{\mu})$,

for $x\geq 0$, with $\underline{\tilde{W}}(0)=\underline{0}$.

d. The quantities $W_j^*(0)$ and $\tilde{W}_j^*(0)$ give respectively the probability that, in the stationary queue, a customer will enter service with the Q-process in the state j and will depart the system with the Q-process in the state j.

These quantities can be used in specific examples to obtain measures of the amount of spill-over from a rush-hour into the subsequent periods of lower traffic.

VI. Some Applications and Problems for Further Investigation

A. Rush-hour Phenomena

In a simple description of an alternating sequence of rush-hours and quieter periods, we construct the Q-matrix as follows.

Let $F_1(\cdot)$ and $F_2(\cdot)$ be PH-distributions on $(0,\infty)$ with representations $(\underline{\alpha}_1,T_1)$ and $(\underline{\alpha}_2,T_2)$ respectively and with means $\kappa_1=-\underline{\alpha}_1 T_1^{-1}\underline{e}$ and $\kappa_2=-\underline{\alpha}_2 T_2^{-1}\underline{e}$. We may assume without loss of generality that the matrices $T_1+T_1^\circ\Delta(\underline{\alpha}_1)$ and $T_2+T_2^\circ\Delta(\underline{\alpha}_2)$, of orders m_1 and m_2 respectively, are irreducible. As usual in discussions of PH-distributions T_1° and T_2° are matrices with identical columns given by the vectors $\underline{T}_1^\circ=-T_1\underline{e}$ and $\underline{T}_2^\circ=-T_2\underline{e}$, respectively. The stationary probability vectors of $T_1+T_1^\circ\Delta(\underline{\alpha}_1)$ and $T_2+T_2^\circ\Delta(\underline{\alpha}_2)$ are respectively denoted by $\underline{\pi}_1$ and $\underline{\pi}_2$ [4].

There is now a convenient way of formalizing the alternating renewal process with underlying distributions $F_1(\cdot)$ and $F_2(\cdot)$. We form the Q-matrix

$$(48)\quad Q=\begin{vmatrix} T_1 & T_1^\circ\Delta(\underline{\alpha}_2) \\ T_2^\circ\Delta(\underline{\alpha}_1) & T_2 \end{vmatrix}.$$

With a slight abuse of notation, T_1° is here an $m_1 \times m_2$ matrix with m_2 identical columns given by T_1°. Similarly for T_2°.

The Q-matrix now defines a Markov chain with $m=m_1+m_2$ states. If the chain is in any one of the states $1,\dots,m_1$, an interval of the first type is "in course" in the alternating renewal process. With the chain in one of the states $m_1+1,\dots,m_1+m_2$, the alternating renewal process is in an interval of type 2.

It is elementary to verify that the stationary probability vector $\underline{\pi}$ of Q is then given by

$$(49)\quad \underline{\pi}=(\kappa_1'\underline{\pi}_1,\ \kappa_2'\underline{\pi}_2),$$

where $\kappa_1'=\kappa_1(\kappa_1+\kappa_2)^{-1}$, $\kappa_2'=1-\kappa_1'$.

We can now model a rush hour, by assuming e.g. that λ_i is large for $1\le i\le m_1$ and small for $m_1+1\le i\le m_1+m_2$. The parameters μ_i can either be independent of i or can be chosen in some judicious way. This leads to an interesting problem in non-linear optimization, which we formulate next.

B. Rush-hour Control

There is a whole class of interesting nonlinear optimization problems associated with the choice of the service rates μ_i, $1\le i\le m$, for the model described above. We may e.g. endeavor to choose the rates μ_i, subject to certain cost constraints, so that the conditional mean queue lengths

(50) $\quad \pi_j^{-1}[\underline{\pi}(I-R)^{-1}]_j, \qquad 1 \le j \le m,$

vary only little with j. This would be one of many ways of smoothing the queue.

An interesting partial result arises from the tractable special case, noted by Yechiali [9]. In our notation and in a slightly refined form, we obtain the following.

Theorem 6

In the particular case, where $\lambda_j=\rho\mu_j$, for $1 \le j \le m$, with $\rho<1$, the equation (31) may be written as

(51) $\quad R=R^2D+\rho D,$

with $D=\Delta(\underline{\mu})[(1+\rho)\Delta(\underline{\mu})-Q]^{-1}$.

The matrix R is then given explicitly by

(52) $\quad R=\frac{1}{2}\sum_{\nu=1}^{\infty}\binom{2\nu}{\nu}\rho^{\nu}D^{2\nu-1}.$

The matrix D satisfies $\underline{\pi}D=(1+\rho)^{-1}\underline{\pi}$, so that (52) implies that $\underline{\pi}R=\rho\underline{\pi}$.

The invariant probability vector $\underline{x}$ of Q^* is then given by

(53) $\quad \underline{x}_k=(1-\rho)\rho^k\underline{\pi}, \qquad$ for $k \ge 0$.

Proof

In this case, one sees by direct substitution that the equation (51) has the same formal solution as the scalar equation $r=r^2d+\rho d$. The series (52) is the matrix analogue of

(54) $\quad r=\frac{1}{2d}\left[1-(1-4\rho d^2)^{\frac{1}{2}}\right]=\frac{1}{2}\sum_{\nu=1}^{\infty}\binom{2\nu}{\nu}\rho^{\nu}d^{2\nu-1}.$

The remaining statements are easily verified.

The qualitative interpretation of Theorem 6 is clear. If the server can produce a service rate $\mu_j = \rho^{-1}\lambda_j$, whenever the arrival rate is λ_j, the stationary queue length distribution will be independent of the state of the Q-process. This ideal smoothing of the queue may however be infeasible in practice. The server may not be able to serve at rates higher than a given value of μ, or cost considerations may make such a high flexibility in the service rate prohibitive.

We do not pursue these topics here. It is important to emphasize however that there is no hope of obtaining tractable analytic solutions for this type of problem in view of the complicated nonlinear dependence through R of the quantities in (50) on the parameters of the problem. A combination of computational experience and techniques from nonlinear optimization on the other hand appears to be promising and will be discussed elsewhere.

C. Interruptions of Arrivals or Services

By setting some of the parameters λ_i and μ_i equal to zero, we can model interruptions of arrivals or services during random intervals of time.

In the main body of the paper, we have assumed that all λ_i and μ_i are positive in order to avoid consideration of particular cases. This assumption can frequently be relaxed in an obvious manner for each theorem. Provided that at least one of the parameters λ_i or μ_j is positive, the matrix A_1^{-1} remains strictly negative. When some of the arrival or service parameters are zero, some of the matrices B_0, B_2, C_0 and C_2 acquire rows or columns which are identically zero and some of the statements

regarding the matrices G and R need to be modified.

For purposes of illustration, we consider the case of service interruptions. We assume that $\lambda_j>0$, for $1\le j\le m$ and $\mu_j>0$, for $1\le j\le m_1$, $\mu_j=0$, for $m_1+1\le j\le m$, with $1\le m_1<m$. In this case the matrix B_0 remains strictly positive and Theorem 5 continues to hold as stated. The statements regarding the busy period require changes, since no busy period can now end when the Q-process is in one of the states $m_1+1,\ldots,m$. In terms of the equation (7), we see that the columns labeled $m_1+1,\ldots,m$ of C_0 are now identically zero. We see that this is also the case for the matrix $G^*(z,s)$ and therefore also G.

A complete discussion of the busy period requires that we show that Equation (6) has a unique solution with the first m_1 columns strictly positive and the other columns equal to zero. The proof of this and of the corresponding moment formulas of Theorem 2 is fully analogous to the irreducible case [5], but requires more tedious steps as we need to partition the matrix $G^*(z,s)$ into the form

$$\begin{vmatrix} G_1^*(z,s) & 0 \\ G_2^*(z,s) & 0 \end{vmatrix},$$

where $G_1^*(z,s)$ is $m_1\times m_2$ and $G_2^*(z,s)$ is $(m-m_1)\times m_2$.

If also some λ_j are zero, the matrix R will have the corresponding rows equal to zero. The form of the vector $\underline{x}$, given in Theorem 5 remains valid, but the proof of the existence and uniqueness of R requires greater care and a consideration of cases.

D. Some Comments on Numerical Computations

Even in the case m=2, the matrix R cannot be obtained in an explicit form, but its numerical computation is straightforward. Successive substitutions in Equation (31) exhibits very rapid convergence, except for cases where ρ is close to one. Computations for problems with m as large as one hundred are entirely feasible and stable.

The approach to numerical computations, described in [9], should however be applied with caution as it involves the computation of the roots of a polynomial equation in the unit interval. Knowledge of these roots permits the computation of the vector $\underline{x}_0$, and the vectors $\underline{x}_k$, $k \geq 1$, can then be recursively computed. If the roots, discussed by Yechiali, are close together, however the computation of $\underline{x}_0$ and therefore of the vectors $\underline{x}_k$, $k \geq 1$, is likely to be of doubtful accuracy.

References

[1] Bellman, R. (1960)
Introduction to Matrix Analysis
McGraw-Hill Book Company, New York

[2] Kingman, J. F. C. (1961)
A Convexity Property of Positive Matrices
Quart. J. Math, 12, 283-284

[3] Neuts, M. F. (1973)
The Markov Renewal Branching Process
Proceedings of a Conference on Mathematical Methods in the Theory of Queues, Kalamazoo, MI, Springer-Verlag, 1-21.

[4] Neuts, M. F. (1975)
Probability Distributions of Phase Type
in--Liber Amicorum Professor Emeritus H. Florin--, Department of Mathematics, University of Louvain, Belgium, 173-206

[5] Neuts, M. F. (1976)
Moment Formulas for the Markov Renewal Branching Process
Adv. Appl. Prob., 8, 690-711

[6] Neuts, M. F. (1978)
Markov Chains, with Applications in Queueing Theory, which have a Matrix-geometric Invariant Probability Vector
Adv. Appl. Prob., 10, (forthcoming)

[7] Purdue, P. (1974)
The M/M/1 Queue in a Markovian Environment
Opns. Res., 22, 562-569

[8] Yechiali, U. and Naor, P. (1971)
Queueing Problems with Heterogeneous Arrivals and Service
Opns. Res., 19, 722-734

[9] Yechiali, U. (1973)
A Queueing-Type Birth-and-Death Process defined as a Continuous-time Markov Chain
Opns. Res., 21, 604-509

[10] Wallace, V. (1969)
The Solution of Quasi Birth and Death Processes arising from Multiple Access Computer Systems
Ph.D. Thesis, Systems Engineering Lab., Univ. of Michigan, Ann Arbor, MI, Tech. Report No. 07742-6-T

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 18 19 AFOSR-TR-77-1233	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER 14 TR-77/11
4. TITLE (and Subtitle) 6 THE M/M/1 QUEUE WITH RANDOMLY VARYING ARRIVAL AND SERVICE RATES		5. TYPE OF REPORT & PERIOD COVERED 9 Technical rept. Interim
		6. PERFORMING ORG. REPORT NUMBER Tech Rpt No 77/11
7. AUTHOR(s) 10 Marcel F./Neuts		8. CONTRACT OR GRANT NUMBER(s) 15 AFOSR-77-3236
9. PERFORMING ORGANIZATION NAME AND ADDRESS University of Delaware Dept of Statistics & Computer Science Newark, DE 19711		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61102F 2304/A5
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research/NM Bolling AFB, DC 20332		12. REPORT DATE 11 July 1977
		13. NUMBER OF PAGES 22 12 25p.
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) 16 2304 17 A5		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

M/M/1 queue, random environment, queue length, waiting time, quasi-birth-and-death processes, Markov chains, nonlinear equations, computational probability

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

~~We study~~ are studied computationally feasible solutions for a number of problems related to a M/M/1 queue in which the arrival and service rates vary according to the state of an underlying Markov chain.

Our results may be used to mode the effect of rush-hour phenomena or other extraneous fluctuations on the characteristics of an M/M/1 queue.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED 391807 S/C

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)